FEATURES FOR AUTOMATED QUALITY ASSESSMENT OF DIGITALLY TRANSMITTED VIDEO

Stephen Wolf*

ABSTRACT

This report describes an automated method of video quality assessment based on extraction and classification of features from sampled input and output video. The first subsystem of the automated video quality measurement system is the feature extraction subsystem. Features are extracted from the sampled video that quantify many of the distortions present in modern digital compression and transmission systems. The feature measurements may then be injected into a quality classification subsystem which will determine the overall quality rating of the video. This report discusses the first subsystem of the automated video quality assessment system, namely the feature extraction subsystem. The measurement techniques used to extract a number of useful features are discussed in detail. Results are presented using sampled video teleconferencing data that contained common video compression artifacts.

Key words: American National Standards; ANSI; feature extraction; image processing; video quality; video teleconferencing

1. INTRODUCTION

As the world prepares to enter the age of digitally transmitted video services such as video teleconferencing/video telephony (VTC/VT), digital television, wideband integrated services digital networks (ISDN), high resolution graphics transmission, and high definition television (HDTV), new quality assessment techniques are needed. Traditional techniques for estimating video quality degradation during transmission have been based on analog measures of the transmission signal. These parameters are not adequate for assessing video quality when images are impaired by the many new types of distortions introduced by the modern digital transmission systems given above. In such cases, the video transmission quality is often a function of the type of imagery being transmitted (line drawings, natural scenes, etc.). Since the information normally has been compressed, small transmission errors due to channel impairments can have significant effects on the received video quality. As a result, viewing panels have been used to evaluate these modern

* The author is with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80303.

distortion effects on video quality. Unfortunately, this approach is time consuming, expensive, and requires special care to prevent wide variations between tests. CCIR Recommendation 500-3 (1986) and Report 405-5 (1986) discuss in detail the methodology for conducting subjective assessment of the quality of television pictures.

New, objective measures of video transmission quality are needed by standards organizations, end users, and providers of advanced video services. Benefits would include impartial, reliable, repeatable, and cost effective measures of video and image transmission system performance and increased competition among providers as well as a better capability of procurers and standards organizations to specify and evaluate new systems.

1.1 Background

Extensive studies have been performed in recent years regarding quality assessment of video pictures. Most of the work falls into one of the following three groups:

1. Subjective quality assessment of still pictures or motion video.

2. Objective quality assessment of video components or systems based on output responses to injected test waveforms or patterns. The objective measurements are sometimes modified to account for characteristics of the human visual system.

3. Objective quality assessment of still pictures or motion video based on extraction of features directly from the video picture. The original (undistorted) picture is usually available for comparison. Since digital sampling of the video is performed, the objective measurements are sometimes modified to account for the effects of the video display device. In addition, characteristics of the human visual system are sometimes incorporated so that the objective measurements correspond more closely to the subjective rating.

CCIR Report 313-6 (1986) provides an extensive bibliography regarding assessment of the quality of television pictures. Nearly all of publications listed in the report deal with the subjective quality assessment described in group (1) above. CCIR Recommendation 567-2 (1986) describes a set of objective measures which fall into group (2) above. CCIR Recommendation 654 (1986) defines relationships between the objective measurements and subjective picture quality, assuming that only one distortion type is present. The works of Biberman (1973), Higgins (1977), Task (1978, 1979), Carlson and Cohen (1980), and Barten (1987, 1988) also fall into group (2) above, since the quality measures are a function of the frequency responses (test waveforms are sinusoids) of the video and human vision systems. Meiseles (1988) has proposed a measurement of dynamic resolution based on rotating test patterns. Group (3) above includes the work of Mannos and Sakrison (1974), Sakrison (1977), Limb (1979), Pearson (1980), Toit and Lourens (1988), Ohtsuka et al. (1988), Miyahara (1988), and Tomich et al. (1989). Here, quality measures are normally developed as a weighted error of the distorted image relative to the original image.

The objective techniques of group (3) above are most applicable to video scenes which have undergone digital compression and transmission. Performance of image compression algorithms are a function of the type of imagery which is being compressed. A compression algorithm designed to perform well on one type of imagery, say natural scenes, may perform poorly on another type of imagery, like line drawings. In addition, the effects of transmission channel impairments (such as bit errors) must be determined by examining the resultant decoded or uncompressed image. Thus, video quality measurements based upon injected test signals, such as the techniques in group (2), could yield objective quality ratings that differ substantially from the subjective quality ratings. For an overview of image data compression, the reader is referred to Nesenbergs (1989).

Very little work in group (3) has been performed on video scenes that contain motion. Even recent papers which propose techniques in group (3) for motion video (Miyahara, 1988, Ohtsuka et al., 1988) do not evaluate their techniques using motion video. In practice, alignment of undistorted video and distorted video (from a wide class of video compression systems) requires careful consideration. Automated techniques for performing proper alignment of undistorted and distorted video will be discussed in detail later in this report.

1.2 Automated Video Quality Measurement System Overview

This report discusses a method for objectively measuring video quality based on feature extraction from digitized video imagery and classification techniques. Figure 1 gives an overview of the automated video quality measurement system. The computer-based approach extracts objective video quality features directly from captured video images. Video quality features extracted from the sampled imagery are chosen to be sensitive to user applications, video compression artifacts, and the effects of modern transmission channel impairments. In this report, a candidate set of features that quantify the presence of video compression artifacts has been developed by the author and his associates. Certain desirable properties of features, to be covered later in this report, guided this initial feature development and selection process. The objectively measured features are interpreted by a quality classification system to produce an overall quality rating comparable to that provided subjectively by a panel of viewers. Subjectively rated video data and psychological results from studies on human perception of video quality are used to assist in the design of the feature extraction and quality classification subsystems. In addition, not shown in Figure 1, certain a priori knowledge may be input to the feature extraction and quality classification subsystems to improve their performance. Examples of a priori control parameters include characterization of the display device which will be used to view the video, the viewing distance, or the type of video service.

The primary goal of the approach is to obtain an objective assessment of video quality that emulates the subjective rating. The goal is accomplished by selecting a set of features measured from the video imagery which correlates well with artifacts noticeable to the viewer, and by incorporating statistical and psychological results obtained from subjective evaluation of video imagery. The candidate set of features will be extracted from subjectively rated video imagery that exhibits a wide range of distortions. Then pattern recognition and classification techniques will be applied to determine the mapping of these objectively measured features into subjective quality space (as in Figure 1). Through application of pattern recognition and classification techniques, some of the features in the candidate set may prove to be redundant or ineffective in determining video quality. Hence, these

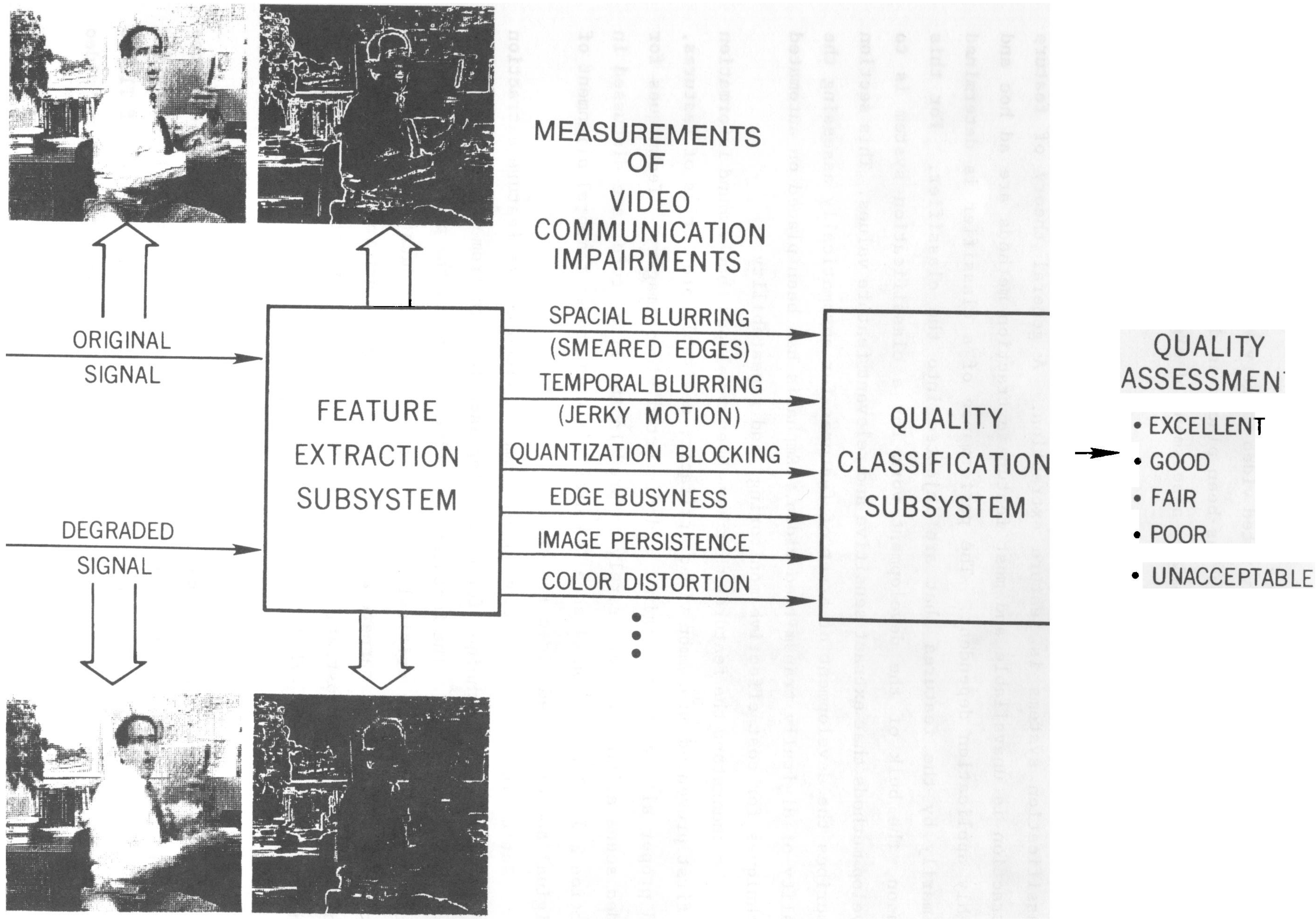

Figure 1. Automated video quality measurement system.

redundant or ineffective features may be discarded. Since subjectively rated video imagery was unavailable at the time of writing this report, emphasis has been placed on development of a candidate set of features for automated quality assessment of digitally transmitted video.

2. DESCRIPTION OF FEATURES

The most difficult process in virtually all pattern recognition and classification systems is feature extraction. A general theory of feature extraction is unavailable and most feature extraction methods are ad hoc and highly application dependent. The performance of a classifier is determined primarily by the features that are injected into the classifier. For this reason, the bulk of the development work for a classification system is to develop methods that extract sensitive and relevant feature values. This section describes the development of a set of features for automatically assessing the quality of digitally transmitted video. Emphasis has been placed on automated techniques for cost effective monitoring, and repeatability.

To understand the features that have been developed, background information is first presented on common video artifacts, desirable properties of features, and proper alignment of original and distorted video imagery. Techniques for video scene alignment, very rarely covered in the literature, are discussed in section 2.3. Calculation of some features requires proper temporal alignment of original and distorted video imagery.

Rationale for preconditioning the sampled video before feature extraction is discussed. The technique for extracting each feature from the sampled video is described in detail. The features objectively quantify the presence of common video artifacts. Of critical concern here is the computational time of a particular feature. Alternate algorithms are presented that reduce this cost of computation. For illustrative purposes, each feature extraction technique is demonstrated using VTC/VT data.

2.1 Common Video Compression Artifacts

The American National Standards Institute, Accredited Standards Committee T1, Working Group T1Q1.5 is drafting interface performance specifications for digital VTC/VT and digital television. The VTC/VT